



RESEARCH DEPARTMENT



REPORT

HDTV Motion adaptive bandwidth reduction using DATV

R. Storey, B.Sc., C.Eng., M.I.E.E.

HDTV MOTION ADAPTIVE BANDWIDTH REDUCTION USING DATV

R. Storey B.Sc., C.Eng., M.I.E.E.

Summary

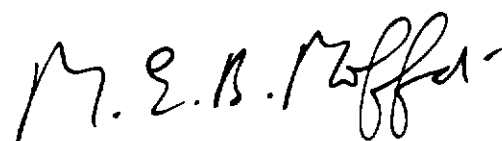
High Definition Television brings with it the problems of an equally high signal bandwidth. This bandwidth will cause problems for all methods of signal dissemination, so some form of bandwidth reduction is required.

Some years ago the BBC successfully developed sub-Nyquist sampling bandwidth reduction techniques. More recently, some promising bandwidth reduction schemes using adaptive filtering followed by sub-sampling have been proposed by others. This Report presents the results of an investigation into the likely performance of such schemes when fully developed. It appears that some of the inevitable impairments caused by bandwidth reduction, although not initially significant, will become more noticeable as the performance of sources and displays improves.

A logical progression is proposed which, by way of a transmitted digital control channel, allows the performance of a 4:1 bandwidth reduced television system to be improved through a series of compatible steps. This Digitally Assisted Television (DATV) progression begins with an interlaced system and finishes with a system which offers all the advantages of a sequential source and display.

DATV is applicable compatibly to conventional and extended definition television systems as well as to HDTV.

Issued under the Authority of



Research Department, Engineering Division
BRITISH BROADCASTING CORPORATION

June, 1986
(PH-280)

Head of Research Department

This Report may not be reproduced in any form
without the written permission of the British
Broadcasting Corporation.

It uses SI units in accordance with B.S. document
PD 5686.

HDTV MOTION ADAPTIVE BANDWIDTH REDUCTION USING DATV

R. Storey, B.Sc., C.Eng., M.I.E.E.

Section	Title	Page
	Summary	Title Page
1.	Introduction	1
2	Anticipated performance	1
3.	The experimental system	1
	3.1 Experimental results	2
4.	Motion vector detection	4
5.	The use of Digitally Assisted Television to achieve a progression in scanning standards	5
6.	Conclusions	7
7.	Acknowledgements	7
8.	References	7

© BBC 2006. All rights reserved. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC Research & Development except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at <http://www.bbc.co.uk/rd/pubs/> for a copy of this document.

HDTV MOTION ADAPTIVE BANDWIDTH REDUCTION USING DATV

R. Storey, B.Sc., C.Eng., M.I.E.E.

1. INTRODUCTION

The high transmission bandwidth required for High Definition Television (HDTV) will inevitably cause problems not only for terrestrial and satellite broadcasting but also for signal dissemination by other media such as videotape, videodisc and cable. Some form of bandwidth reduction is clearly required in order to overcome these difficulties.

Methods of bandwidth reduction have been described which use sub-Nyquist sampling with pre-filtering in one, two and three dimensions^{1,2}. More recently motion adaptive pre-filtering techniques have been described^{3,4}. These systems are in general based upon the removal, by filtering, of image frequency components that are assumed to be of little use to the eye. The filtered signal has a much reduced bandwidth and can be re-sampled at a lower rate for transmission.

It has been widely assumed in past work that the eye cannot make use of high spatial frequencies if they are moving, and this is indeed true for a fixed gazing point⁵. For normal television pictures however, the eye will generally attempt to follow moving areas of interest either by continuous motion for low speeds or by saccadic motion for high speeds⁶. The eye's spatial detail requirements in moving areas may well be reduced but certainly not, in the case of uniform well correlated motion, by the large factors assumed in most of the literature.

In order to discover the amount of bandwidth reduction attainable by these techniques, evaluate possible impairments and discover any operational problems, it was decided to build an experimental system.

A deliberate decision was made at the outset to build a system operating at an accessible line rate. By choosing a value of 15,625 lines/sec, the same as is used for present day 625 line television, it was possible to guarantee that the signals were originated and displayed by mature hardware, capable of delivering its full theoretical resolution. The source and display would not therefore be limiting factors so the performance attainable, after appropriate scaling, would reflect that of a fully matured HDTV bandwidth reduction system.

2. ANTICIPATED PERFORMANCE

Substantial bandwidth reduction may be achieved, in stationary areas, by mild diagonal spatial filtering followed by spreading the transmission over four fields. A complete, highly detailed, stationary image

will therefore take the same number of field periods to accumulate. This is not in itself a serious limitation; a computer simulation has shown that, for a new scene, a progressive increase in detail content spread over four fields is barely perceptible.

Moving areas must however be treated differently, with priority given to smooth portrayal of motion. Simple protraction of the time domain cannot be tolerated so bandwidth reduction must be achieved by a further reduction in spatial resolution. Thus an adaptive system, which treats moving and stationary areas differently, is required.

The performance of such a system might therefore be expected to be good for stationary images, and this was shown to be the case by computer simulation. As an object in the scene begins to move however, the system changes to the mode offering smooth motion portrayal at the expense of spatial resolution. The motion speed at which this change should best occur, and the visibility of the resulting detail loss when compared to a fully resolved background, were significant unknowns. Other uncertainties were the feasibility of keeping the decoder and encoder operating in unison, the nature of artefacts to be found at transitions from one transmission mode to another and the effects of transmission channel noise on bandwidth reduced signals.

3. THE EXPERIMENTAL SYSTEM

The first stage of the work was to build an experimental system whose functional block diagram is shown in Fig. 1. The system was not constrained to be interlaced nor to a fixed field rate, these parameters being fully programmable. In common with previous work⁴ there were two pre-filters whose purpose was to reduce the signal bandwidth in the vertical-temporal domain for stationary areas and in the spatial domain for moving areas. The reduced bandwidth signals were re-sampled at a lower rate to form two alternative transmission signals. Note that neither of these sub-sampled signals was suitable for direct display at the receiver; interpolation was necessary to reconstruct a displayable signal.

Both the manner in which the most appropriate transmission signal was chosen and the communication of that choice to the receiver, differed significantly from the prior art. Each subsampled signal was restored to a displayable form in the encoder, using interpolators identical to those in the remote decoder. In this manner, the encoder had access to the signal that the decoder would generate if it were to use the correct type of interpolator. The fidelity of the two alternative filtering schemes was measured by subtracting each reconstructed

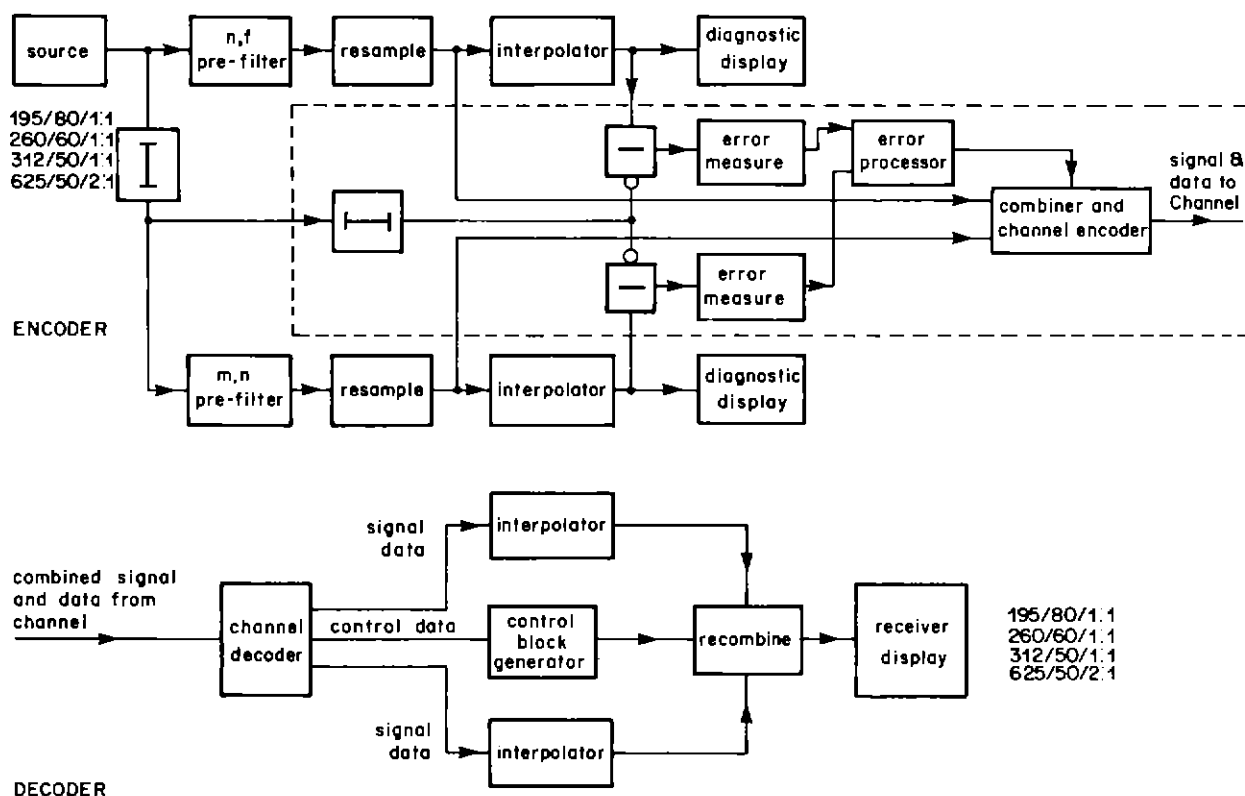


Fig. 1 - Experimental arrangement for investigation of motion adaptive bandwidth reduction

signal from the original. These differences, which were effectively coding errors, were accumulated over a spatial aperture or block. The signal which gave the smallest coding error within each block was chosen for transmission, and was switched into the transmission channel along with a digital control signal which told the decoder which type of reconstruction to use.

In this approach to HDTV transmission, all coding decisions are taken at the encoder where the original undistorted signal is available for comparison. The transmission link is configured as an analogue image signal augmented by a rugged digital control signal; we have called this method Digitally Assisted Television (DATV). The decoder is simply told how to reconstruct a displayable signal rather than having to take decisions itself, based on a transmitted signal with a much reduced information content and a much poorer signal-to-noise ratio. In this manner, the system performance is enhanced at the same time as the cost and complexity of the receiver are reduced.

3.1 Experimental Results

To date, pre-filters and interpolators have been designed for interlaced source and display only, and for bandwidth reductions of 2:1 and 4:1. The scope for effective pre-filtering is, however, severely restricted by the use of an interlaced source. A conventional interlaced camera for example, can resolve only about 0.65 of its potential vertical resolution if its integration time is to be

held down to an acceptable value (nominally 1/50 sec.). This constraint requires that the entire target should be discharged on each field, which is achieved in practice by enlarging the scanning spot to cover not only the current line, but also the region occupied by adjacent interlaced lines. The output of the camera is therefore reduced for all high spatial frequencies. Horizontal resolution can be largely restored by the use of aperture correction. Any attempt to restore vertical resolution at frequencies approaching $N/2$, where N = the number of picture lines, will however be less successful. This is because high vertical frequencies in the original scene can give rise to the same signal components as moving low vertical frequencies, rendering the two indistinguishable. It is therefore difficult, if not impossible, to restore full vertical resolution without causing motion impairment. In short, the use of an interlaced source means that the available alias-free vertical definition in stationary areas is restricted to just over half of the transmission channel capacity, and is only slightly better than the transmissible vertical resolution for motion.

By a similar argument, an interlaced display cannot reproduce vertical frequencies much higher than $0.65N/2$ cycles/picture height, because the eye prefers to interpret these as vertical twitter at half the field rate rather than stationary vertical detail.

For a 4:1 bandwidth reduction, the most visible effect in moving areas is therefore a halving of horizontal resolution; this effect is clearly illustrated in Figs. 2(a), (b) and (c). Fig. 2(a) shows the resolution available, before



Fig. 2a. Available detail before bandwidth reduction.



Fig. 2b. Loss of detail in the image of the moving person caused by bandwidth reduction.



Fig. 2c. Momentary reversion to the highly detailed stationary mode, during a halt in motion.

bandwidth reduction, when an interlaced camera and display are used. Fig. 2(b) shows a similar image after bandwidth reduction. The car and background are stationary and have therefore suffered very little loss of detail. The person behind the car however, is moving with respect to the camera; his image is transmitted in the motion mode and has consequently suffered a loss of resolution. It is worth noting that the bandwidth reduced image is remarkably free from switching artefacts at the boundary between stationary and moving areas. This level of dynamic performance should be maintained in the varying noise environment found in a real satellite channel, since the digital motion data can be heavily protected.

Unfortunately, the motion speed beyond which the stationary transmission mode becomes unacceptable has proved to be disappointingly low, so for real pictures the system spends most of its time in the low detail, motion mode. The loss of detail is not quite as distinct for unpredictable or poorly correlated motion and would probably be quite acceptable in areas containing such motion.

This loss of resolution on moving objects is highlighted when they halt momentarily or change direction; for a brief instant their resolution reverts to that of the highly detailed stationary mode. This effect is shown in Fig. 2(c), which is again a similar scene to that of Figs. 2(a) and (b), but caught at an instant when the person's head was stationary with respect to the camera. This effect is accentuated by erratic but predictable motion, such as is found in a close-up of a person talking. Under such conditions the intermittent loss of moving detail shown in Figs. 2(b) and (c) proves to be immediately obvious and quite objectionable.

It is worth noting that a 4:1 bandwidth reduction system operating at a 'high definition' standard would not in its early days show quite as severe a loss of moving detail as revealed by our experiments. Current high definition cameras are at the beginning of their development life and do not have as high a resolution as a fully matured product. As camera performance improves, the added vertical resolution will become an embarrassment since it will simply cause more inter-line twitter, and the added horizontal resolution will throw the loss of detail in moving areas into sharp relief. Such a system would not be 'future proof'.

The coding fidelity measurement and control channel techniques have proved to be very effective. The block size should be comparable to the aperture of the pre-filters and interpolators to minimise the appearance of artefacts at transitions from one transmission mode to another, but not so large as to produce boundaries identifiable from a normal viewing distance. A vertical and horizontal dimension of about 1/100th of a picture height proves to be a good compromise and yields

manageable data rates, in the region of 1-2 Mbits/sec, for the control channel.

4. MOTION VECTOR DETECTION

The problem of detail loss in well correlated moving areas can be solved, in theory, by detecting vectors describing the direction and speed of the predominant motion in a scene. A limited number of vectors, say two or three, would suffice to identify the salient areas of motion. These motion vectors are assigned to the appropriate parts of the scene and used to displace the reference frame of the high spatial detail pre-filter, in such a way that moving areas appear to be stationary. They can then be transmitted in their correct spatial position and with full spatial detail.

The decoder is told the values of motion vectors for the following field during vertical blanking, and which areas belong to which motion vector continuously, via the control data channel. It then applies complementary displacements to its stationary interpolator's reference frame to reconstruct highly detailed moving areas. Displacing pre-filter and interpolator reference frames rather than the bandwidth compressed signal permits a receiver without motion compensation to use the motion compensated signal, thereby preserving compatibility. In such a receiver, the motion compensated areas are decoded using a modified motion mode, thus allowing them to be reproduced in the correct position but with the previously noted poorer spatial resolution.

Effective motion vector detection has been achieved in a computer study by calculating the correlation surfaces for a series of time displaced images containing complex motion. The two basal axes of the surfaces are horizontal and vertical displacement and the third axis describes the degree of correlation. Each area that has suffered a displacement between two adjacent images gives rise to a peak in the correlation surface centred upon the co-ordinates of its displacement. The surface is interrogated to find the exact positions of the largest correlation peaks which then become the salient motion vectors between those images.

The next operation is to assign motion vectors to their corresponding image areas. This is achieved by shifting the entire source image by each of the motion vectors in turn and forming an error surface whose basal axes are horizontal and vertical position. Areas of the image belonging to each vector will give rise to local nulls in the corresponding error surface and can then be assigned to that vector. An error surface based on a zero displacement will, of course, identify stationary areas, so the outcome of the matching process will be either stationary areas, areas moving with identifiable motion vectors or areas which are changing in an unidentified manner.

The assignment of motion vectors to their corresponding image areas could well be combined with

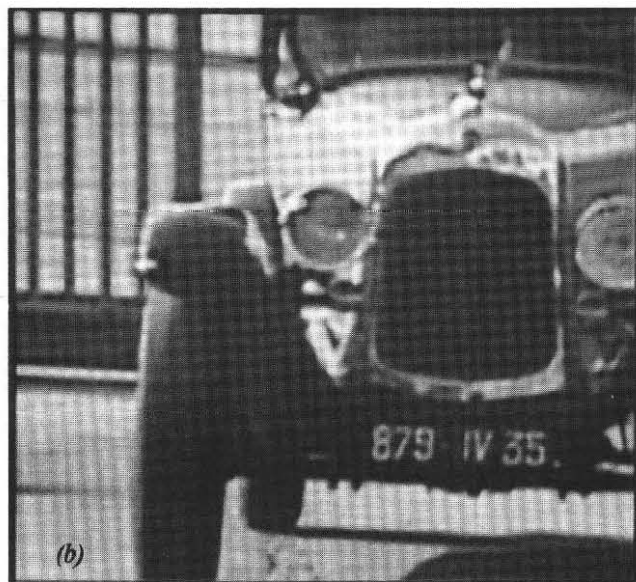
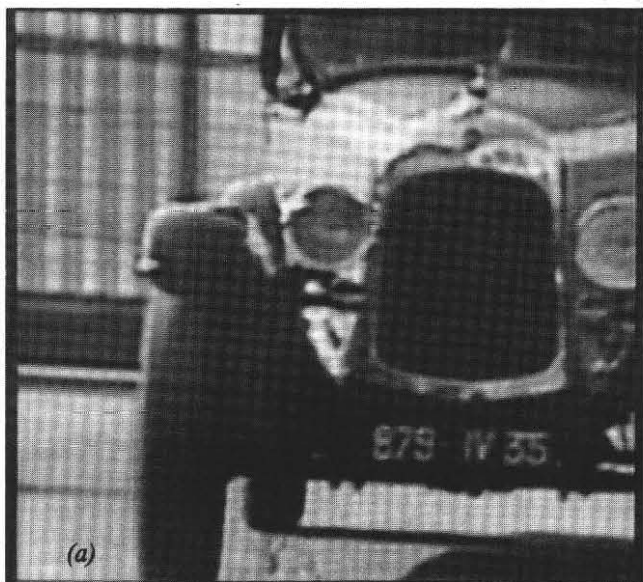


Fig. 3. Interpolation of an image from preceding and following television fields (a) without and (b) with motion vector compensation.

the coding fidelity measurement to good effect. It remains to be discovered whether a separate assignment of motion vectors followed by coding fidelity measurement, or a combined approach in which the coding fidelity of each motion vector is tested independently, provides the best solution.

An indication of the potential of localised motion vector detection is shown in Fig. 3. The source sequence from which these two images were derived consists of a car moving into the foreground and turning at the same time. The background and gate are also moving, so the sequence contains differential motion, parallax and rotation.

The image shown in Fig. 3(a) was derived from the preceding and succeeding television fields by simple first order temporal interpolation. Positional errors can be clearly seen as a spreading in the bars of the moving gate and as a loss of detail in the car. The image shown in Fig. 3(b) is derived from the same two source fields using motion vector detection and correction. Positional errors in the gate do not now occur and detail has been restored in the moving car.

The next stage of the work is to construct real-time hardware to perform the motion vector detection algorithms developed in the computer study and to combine this with the bandwidth reduction equipment already built. This should provide a system capable of transmitting a highly detailed image in stationary and well correlated moving areas. The received image will revert to a reduced detail mode only in poorly correlated moving areas such as those containing rotations or erratic motion, and in areas of newly revealed detail. The transmission system will not be perfect but, if the motion vector

compensation proves effective, its failings should be restricted to those areas in which the eye would be truly unable to perceive them.

5. THE USE OF DIGITALLY ASSISTED TELEVISION TO ACHIEVE A PROGRESSION IN SCANNING STANDARDS

It is widely accepted that a sequential camera would be capable of providing a superior vertical resolution to that available from an interlaced camera having the same number of lines and the same integration time, albeit at the expense of an initially higher signal bandwidth⁷. It has also been shown that successful display of this extra vertical resolution can be achieved using a sequentially scanned display⁸. Indeed the bandwidth reduction systems described in this Report can benefit significantly from the use of a sequential source and display. The required transmission bandwidth would remain unchanged but the available vertical resolution would be increased to a value approaching the theoretical maximum of $N/2$ cycles/picture height, where N is the total number of scanning lines.

Drawing upon the ideas outlined in this Report it is possible to identify a logical progression from an initial wholly interlaced system, through a series of compatible intermediate stages, culminating eventually in a sequentially sourced and displayed system, as shown in Fig. 4.

In the early days of an emerging HDTV system, during the first stage in the progression, sources and displays will probably both be interlaced and will certainly be limiting factors in the available resolution. If transmission is by means of a bandwidth reduction system, an HDTV receiver would need to incorporate a

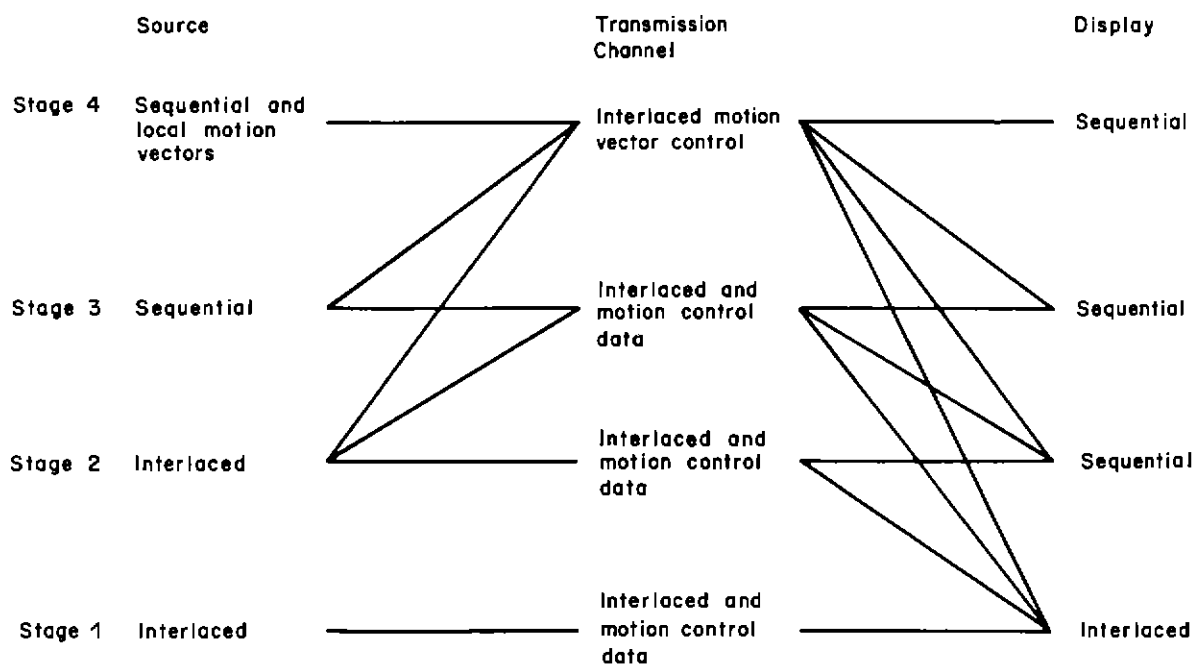


Fig. 4. A logical progression for the development of HDTV.

matching decoder in any case. A little advance planning of the data channel would easily clear the way for the progressive introduction of more developed systems capable of higher resolution. At the same time, a second receiver, possibly with a smaller display, might not require the full resolution and could use the data channel to control a much simplified decoder which treats all the incoming signal as if it were moving.

The second stage of the progression occurs when sequentially scanned displays become available. In this case, the conversion process from an interlaced signal to a sequential display has been shown to require information about the image's motion content⁸. Again this information already exists in the data channel. It should be noted however that the only improvement at this stage would be a reduction in inter-line twitter and a marginal increase in vertical resolution.

A significant increase in vertical resolution would occur at the third stage in the process, when sequentially scanned cameras, which can differentiate between high frequency vertical detail and motion, become available. At about that time, the resolution of the sequential display should have improved to the point at which the extra resolution could actually be realised.

A parallel step, which incorporates motion vector detection, is shown as stage four in the progression. This extends the higher definition mode into well correlated moving areas, leaving only poorly correlated areas and scene changes to be transmitted at a lower definition. The technical complexity required to include motion vector compensation at the decoder is probably comparable to

that required for up conversion to a sequential display. For this reason it is arguable which should come first and it may well be advisable to introduce the two ideas at once.

Although the arguments presented so far have been directed towards the transmission of high definition images, the digital assistance principle can be used to improve the performance of systems having a lower basic definition such as MAC or current PAL and NTSC. A simple motion signal, sent via a digital assistance channel, could make display up-conversion to a higher line number, or to a sequential scan, more effective by helping to differentiate between motion and high frequency vertical detail. The resulting potential increase in vertical detail could be complemented in the horizontal direction by using bandwidth reduction techniques⁹, and here again a reliable motion signal would be invaluable.

The further addition of motion vector information could, with sufficient processing power, permit the interpolation of intermediate fields with correctly positioned moving objects. This would largely remove artefacts such as the combing of moving horizontal detail caused by simple, non-motion corrected, temporal interpolation between adjacent moving fields⁸.

A reliable motion signal could also simplify the implementation of video noise reduction at the receiver by avoiding the need for the complex noise measurement circuits required for a remote motion detector¹⁰. This could be a very useful facility for a satellite receiving system where the size of receiving dish is limited for other reasons.

6. CONCLUSIONS

The performance of a digitally assisted 4:1 bandwidth reduction scheme based on adaptive subsampling has been described and substantiated by experimental evidence. Some potential problems, which are likely to occur within a short time of the introduction of such systems, have been identified.

The principal limitation at the outset is the inability of an interlaced source and display to resolve and subsequently display the full vertical spatial resolution available from the bandwidth reduction system. In addition, the horizontal resolution of sources will also be a limiting factor in the early days. As the performance of sources and displays improves, the additional vertical resolution will manifest itself as interline twitter and the additional horizontal resolution will give rise to an objectionable difference in relative definition between moving and stationary areas.

These problems can be overcome to a large extent by the addition of a sequential source and display combined with localised motion vector detection to extend maximum spatial definition into the majority of moving areas. Such enhancements are consistent with the framework of a Digitally Assisted Television (DATV) system since they can be achieved simply by increasing the bandwidth of the digital control channel.

A logical progression has been described which allows the step by step development of an HDTV system, beginning with a technology-limited interlaced standard, progressing through a number of compatible intermediate steps, and finishing with a sequentially sourced and displayed system. The end product would be capable of producing the full potential static and dynamic resolution, losing definition only in poorly correlated areas. The transmission channel is practicable and remains unaltered throughout.

The remaining small impairments should be a small price to pay in exchange for a system which, with the assistance of DATV, can be developed in a compatible progressive manner, to give all the potential resolution and freedom from display artefacts that are known to be possible with sequential scanning. DATV is applicable compatibly to conventional and extended definition television systems as well as to HDTV.

7. ACKNOWLEDGEMENTS

I would like to thank G.A. Thomas for his work on motion vector detection and T.A. Moore for his work on computer simulation of filters and the effects of a progressive accumulation of image detail.

The source pictures for investigations of motion vector detection were provided by courtesy of CCETT Rennes.

8. REFERENCES

1. DEVEREUX, V.G. and PHILLIPS, G.J. 1974. Bit-rate reduction of digital video signals using differential p.c.m. techniques. IEE Conference Publication No 119, International Broadcasting Convention, September 1974, pp 83-89.
2. OSBORNE, D.W. and STOTT, J.H. 1979. Digital video: comb filtering in sub-Nyquist PAL codecs. BBC Research Department Report No. BBC RD 1979/1.
3. GLENN, W.E. et al. Compatible Transmission of High Definition Television using Bandwidth Reduction. Proceedings of 37th Annual Broadcasting Engineering Conference, National Association of Broadcasters, 1983. pp 341-349.
4. NINOMIYA, Y. et al. 1984. A Single Channel HDTV Broadcast System, The MUSE. NHK Laboratory Note No. 304.
5. WATANABE, A. et al. 1968. Spatial Sine-Wave Responses of the Human Visual System. Vision Res Vol. 8, Pergamon Press. pp 1245-1263.
6. LINDSAY, P.H. and NORMAN, D.A. 1977. Human Information Processing. Academic Press (London).
7. WENDLAND, B. 1983. Extended Definition Television with High Picture Quality. SMPTE Journal, October 1983, pp 1028-1035.
8. ROBERTS, A. 1983. Improved Display of 625 Line Television Pictures. BBC Research Department Report No. BBC RD 1983/1.
9. ARRAGON, J.P. et al. 1986. (Member of Philips Research Organisation.) Transmission d'images de television a haute definition compatible avec le standard D2-MAC/paquet. Acta Electronica 27, 1-2.
10. DREWERY, J.O., STOREY, R., TANTON, N.E. 1984. Video Noise Reduction. BBC Research Department Report No. BBC RD 1984/7.

